

Policy on Archiving Qualitative Data

In addition to its collection of quantitative social science data, the FORS data service solicits, preserves, and disseminates qualitative research data. The following sections detail our current policies and procedures.

1.0 Acquisition: criteria for selection

FORS is a centre of expertise in the social sciences, and so we focus only on data that fall within the realm of the social sciences.

The qualitative data should:

Have further analytic potential. We want to provide data that can bring new insights to the social sciences. Therefore, data should have further analytic potential than the original investigation.

Represent high quality/exemplary studies for qualitative social research. This means that projects should be good to excellent examples of well-designed and conducted qualitative research, and should be of potential use as models for teaching research methods.

The data and documentation should preferably be digital, and could include audio, image, and/or video formats. However, we reserve the right to deny archiving of projects for which digitalisation is needed on a large scale. The projects must satisfy our criteria concerning documentation and anonymisation (see following sections).

Most important is the collaboration with the primary researcher, his/her interest in having the data archived and made accessible to the public, and his/her willingness to be involved in some ways in the documentation, access, and reuse processes.

2.0 Confidentiality, anonymisation, informed consent, and access

FORS relies on a combination of informed consent of study participants, basic anonymisation, restrictive end-user contract conditions, and access eligibility conditions as a way of ensuring the confidentiality of data.

Specifically, preservation of confidentiality of qualitative data is based on the following four measures:

1. Informed consent by respondents, usually at the time of fieldwork, for "use of the collected data for research purposes";
2. Basic anonymisation and removal of identifying information in transcriptions;
3. Access restricted to researchers and students affiliated with a research institution.
4. Contracts that legally bind users to specific terms, including proper use of the data and respect for the confidentiality of individuals.

As applicable, researchers depositing their data should provide information concerning:

- Written informed consent forms;
- A statement on how consent was obtained (e.g. with oral consent);
- Communications with informants relating to confidentiality;
- Anonymisation protocols

2.1 Anonymisation of data

In order to ensure confidentiality, anonymisation of qualitative data is vital to the extent that this is possible and appropriate. Qualitative data to be deposited at the FORS data service should be anonymised *by the researchers* as far as feasible, but the changes should not compromise the quality and usability of the data. Anonymisation of transcriptions should involve removing the most important identifying details, such as proper nouns, addresses, and so on, while replacing this information with pseudonyms and keeping track of all the changes with a table that links original features with pseudonyms.

We can provide guidance to researchers on how to anonymise their data. However, data service staff will *not* be responsible for anonymisation of text files. FORS does not require that audio, image, or video files are anonymised, and is not responsible itself for this. Access will be given to audio, image, and video files only if study participants have consented to use of the data for research purposes.

2.2 Informed Consent for re-use

We doubt that many researchers in Switzerland currently as a matter of practice ask for participant informed consent for archiving and reuse of their data. From an ethical point of view, this is problematic for later archiving of the research data, especially in qualitative research, where participants often reveal sensitive information. On the other hand, it is not always feasible to ask for consent for archiving, depending on the research topic. We strongly recommend therefore that researchers ask their study participants for their consent for "use of the collected information for research purposes", or some variation of this.

While we will not make such a consent a necessary condition for deposit, whether or not such a consent has been granted will affect conditions for data access (see later section on access). Most notably, no access to audio, images, or video will be granted without such participant consent.

2.3 Data access

For now, eligible users are those researchers within Switzerland affiliated with a university or research institution, although researchers from outside of this system can apply for access. Such outside requests will be carefully vetted. Generally, eligible users are those with research or teaching purposes and from research institutions or students enrolled at an institution of higher education.

Access eligibility is a necessary but not sufficient condition. All users must justify their need for the data and must demonstrate a certain level of competence in their written proposals. Requests are judged on a case by case basis.

With respect to access, the following conditions apply:

- For regular research usage, with end-user contract:
 1. access to non-fully anonymised transcriptions, audio/images/video, only with participant informed consent
 2. access to fully anonymised text transcriptions or other anonymised materials

- If reuse is for teaching purposes, with end-user contract:
 1. access to non-fully anonymised transcriptions, audio/images/video files *on a case by case basis*, only with the informed consent of participants
 2. access to fully anonymised text transcriptions or other anonymised materials

In some cases, there may be special access conditions stipulated by the data producers, for example, obtaining their permission, or limiting access to specific types of users. Data will only be released to users with respect to such conditions.

3.0 Study and data documentation ¹

At a minimum, study- and data-level documentation should be sufficient to allow a new user to understand the original research underlying the particular data, and to make appropriate reuse of the data. The producer of the data should agree that the documentation and supporting materials provided to users are sufficient to allow for adequate contextualisation and interpretation of the data. FORS has adopted the following policy regarding elements and forms of documentation for qualitative data:

3.1 Study level

- 1) Documentation standard: DDI2 (Data Documentation Initiative)

We follow the DDI2 international standard for metadata, since this is consistent with existing documentation for quantitative data at FORS. This allows similar study level description of qualitative and quantitative data, and allows for qualitative data discovery using the [FORSbase data catalogue](#), rather than requiring a separate search system. We use a set of DDI2 elements already in place that are suitable for both quantitative and qualitative data, such as:

- title,
- author,
- date of production,
- funding agency,
- bibliographic citation,
- abstract,
- mode of collection,
- method instruments used,
- temporal coverage,
- geographical coverage,
- population,

¹ Some of the details of this document have been drawn from different existing sources, including the Council of European Social Science Data Archives (CESSDA) recommendations and various European data services, including the UK Data Archive.

- sampling/selection procedures,
- key words
- method description.

2) Description of collection methods, and researcher reflections on the fieldwork conducted

We also include information on the following aspects of data collection (as applicable), as well as any notes on the fieldwork:

- collection process,
- coding schemes and categories,
- hardware and software used.

3.2 Data level

Each project deposited to the data service should have a document listing the data files included (for example, interview transcriptions, audio recordings). The file description should include file formats, pertinent characteristics of participants, and any other key information (e.g. language of the interview). The listings should be completed by the data producers, but data service staff can assist if necessary.

Individual transcripts in text documents should include at least²:

- a case identifier,
- a document header giving brief details of the data collection event, including date, place, interviewer name and interviewee details,
- a uniform layout throughout the research project,
- speaker tags indicating the question/answer sequence,
- pseudonyms to anonymise personal identifying information,
- line breaks,
- page numbers.

4.0 Data discovery and distribution

4.1 Catalogue and availability of information

Study metadata are entered into our archive databases, and thus are searchable by way of our [data catalogue](#) in FORSbase. Once identified by way of the catalogue, interested users will sign an end-user contract and be able to download the data together with the necessary documentation.

5.0 Formats

Toward the goals of long-term and high quality data preservation, and ease of deposit for researchers, DARIS has established a set of preferred and accepted data and documentation formats. These take into consideration researcher collection tools and common formats, balanced

² From UKDA's "Managing and Sharing Data: A Best Practice Guide for Researchers"
<http://www.esds.ac.uk/news/publications/managingsharing.pdf>

with the data service needs for standardised data storage, possible format migration, and data dissemination.

Exhibit 1: Preferred and accepted formats

Category	Type of data	Preferred formats	Accepted formats
	Text	Adobe Portable Document Format PDF/A, PDF (*.pdf) MS Word (*.doc, *.docx)	OpenDocument Text (*.odt) Rich Text Format (RTF) (*.rtf) Hypertext Mark-up Language (HTML) (*.html) eXtensible Mark-up Language (XML) according to an appropriate schema (*.xml)
Qualitative or Documentation	Plain text	Unicode, Non-Unicode and ASCII (*.txt)	
	Audio*	Waveform Audio Format (*.wav)	MPEG-1 Audio Layer 3 (*.mp3) Free Lossless Audio Codec (FLAC) Audio Interchange File Format (AIFF) (*.aif, *.aiff)
Qualitative	Pictures: Raster (bitmap) images	TIFF (*.tif, *.tiff)	Adobe Portable Document Format (PDF/A, PDF) (*.pdf) JPEG (*.jpeg, *.jpg) JPEG 2000 (*.jp2, *.jpx) PNG (*.png) GIF (*.gif) BMP (*.bmp)
	Pictures: Vector images	Scalable Vector Graphics SVG (*.svg) Adobe Illustrator (*.ai)	Drawing Interchange File Format DXF (*.dxf) PostScript (*.eps)
	Video	MPEG-4 (*.mpg4)	Moving Picture Experts Group MPEG-2 (*.mpg2) Motion JPEG 2000 (*.mj2)

			QuickTime (*.mov) Lossless AVI (*.avi)

DARIS – list of preferred and accepted file formats for deposit